Retaliation, Resolutions and Remainders: An Argument Against Genuine Moral Dilemmas

By

Daylian Cain

Submitted in partial fulfillment of the requirements for the degree of Master of Arts

at

Dalhousie University

Halifax, Nova Scotia, Canada

September, 1997

TABLE OF CONTENTS

Acknowledgments

I wish to give thanks to the following people:
Julie Barkhouse, Richmond Campbell, Nathan Brett, Micheal Fernandes, Melinda Hogan, Dennis Klymchuk, Jason MacIsaac, Bob Martin, Steve Maitzen, Grant Raynard, Nancy Salay, Peter Schotch, Francesca Wong and a special thanks to Duncan MacIntosh, to whom this paper and my interest in philosophy is dedicated.

I also want to acknowledge a conversation with 'Steve,' a classmate from Yarmouth High School, in Nova Scotia. Steve and I worked in the woods prior to my going to university and it was with him that I recall my earliest and most fierce arguments to the effect that 'want' implies 'can' and does so without qualification, conflict, or dilemma. The sentence that got me started was: 'Daylian, I do not *want* to go fishing; I do go fishing, but I *want* to stay home.' I vehemently insisted that 'want' implies 'do,' and since 'do' implies 'can,' I insisted that 'want' implies 'can.' Much of our argument remains unresolved to this day. He became a fisherman, and I became a philosopher. I hope that someday, he will be able to stay at home.

Abstract

There has been much debate over the idea that the true moral code could oblige incompatible acts and thus allow for genuine moral dilemmas. Against the notion that genuine moral dilemmas exist are arguments that suggest that an action-guiding principle cannot enjoin mutually incompatible acts, for to perform both actions would be impossible and therefore not morally required. Others object that what is required by a set of conflicting principles is not *impossible* to do, but add that ours is a 'dirty' world-- sometimes we are damned if we do and damned if we do not -- so any ethical system which is devoid of moral dilemmas is unrealistically 'neat' and does not account for the feelings of regret we (properly) have when choosing 'the lesser of two evils.'

In this paper, I address these arguments for and against genuine moral dilemmas and then use the Deterrence Dilemma as a case study on which to test my intuitions about these arguments. I get on side with those who argue against moral dilemmas, but the way in which I do so seems to skirt around the controversy over some of the deontological methods of arguing against moral dilemmas. This paper also provides possible objections and innovations to the theories of David Gauthier and of Duncan MacIntosh, at least as these philosophers purport to solve the various paradoxes of rational choice (one of these paradoxes being the Deterrence Dilemma). By the end of this paper, I hope to have tidied away some possible sources of damnation and regret and to have shown that ours is a 'cleaner' world than we first supposed.

Chapter 1: Arguments For and Against Moral Dilemmas

AGAINST:

Sometimes, when faced with what appear to be genuine moral dilemmas, it seems as if we are 'damned if we do and damned if we do not' and all that is left for us to do is to choose the lesser of the two evils. Could it be that the true moral code is 'act-inconsistent,' or requires incompatible acts? If so, there will be times when, no matter what we do, we will fail to comply with at least one moral obligation; and no matter how attentive we are to our goals, we will be unable to avoid feeling regret. Such could be our lot in life if our world is one where genuine moral dilemmas abound. In this paper, I will explore the possibility of avoiding this fate.

One can turn to deontic logic to find a powerful argument against the claim that there are genuine moral dilemmas. Deontic logic originates from the works of Mally (1926) and von Wright (1951) and is the branch of modal logic that deals with obligations and the connections between sentences concerning what one *ought to*, *must*, or *is permitted to* do. The argument from deontic logic against moral dilemmas might go like this: if one ought to perform act A, and ought to perform act B, then, by the 'agglomeration principle,' one ought to perform act 'A&B;' moral dilemmas arise when one's performing of act B is incompatible with one's performing of act A, so B can be thought as simply 'not-A.' The result is that the claim that there are genuine dilemmas suggests that, by agglomeration, one sometimes ought to 'A & not-A;' but, if 'ought' implies 'can,' there is never an obligation to bring about 'A & not-A' (for one

can't); so, by *modus tollens*, one is never in a genuine moral dilemma. To deny this conclusion, one would have to deny the agglomeration principle (as does Williams: 1965), or deny that 'ought' implies 'can' (as does E. J. Lemmon: 1962).

As discussed by Terrance McConnell, Earl Conee, and Bas van Fraassen, there exists a further tension between deontic logic and the claim that there are genuine moral dilemmas. According to a principle of deontic logic, 'the logical consequences of what ought to be, ought to be' (van Fraassen: 1973). So given any act A, since one's performing of A has the logical consequence that one does not perform any act incompatible with one's performing of A, then (by that principle of deontic logic) if one ought to perform A, then it ought to be that no act incompatible with performing A is performed. But the appearance of moral dilemmas is initiated by assertions that one ought to perform A and that one ought to perform some other act incompatible with performing act-A. So we see how yet another principle of deontic logic might run contrary to the claim that genuine moral dilemmas exist.

Other obstacles stand in the way of genuine moral dilemmas. For instance, if we understand realism as making the claim that a statement's truth-value is determined by a mind-independent world, then moral realism ('cognitivism') would also seem to disallow genuine moral dilemmas. The existence of two conflicting but true moral statements would, for the cognitivist, unacceptably require that the mind-independent world which conferred truth upon these incompatible statements contain a contradiction. Even the neophyte to moral theory might have similar misgivings when

presented with the concept of genuine dilemmas. Whenever we find that a certain set of rules leads to a conflict, we often demand that the conflict be resolved and that the rules be revised to avoid further confusion. Regardless of whether one is a cognitivitist, insofar as rules (moral or otherwise) are supposed to guide action, it seems reasonable to demand that such rules do so with clarity and without conflict.

One method of denying the existence of moral dilemmas, which I shall call the 'Conditional-Substitution' method,[1] is to add exception clauses to the so-called 'moral standards' which lead to conflict. If the exception clauses are added properly, the newly conditionalized standards will never conflict. Consider the following example: a code of medical ethics might contain the plausible moral rules that one should respect a patient's autonomy and that one should be non-maleficent. These rules might conflict, say, when respecting the autonomy of a patient who has requested assisted suicide demands that the patient be mercifully killed by the caring physician, but the physician's being non-maleficent precludes acts of killing. If the medical profession reasons that, ultimately, the right course of action in such cases is to respect autonomy, then the profession might revise the code of ethics so as to read 'Always respect the autonomy of your patients and harm your patients *in and only in those exceptional cases* where respecting their autonomy justifies a merciful killing of them.' Those who follow the newly conditionalized code of ethics will perform doctor-assisted suicide, just as the medical profession reasons that one ultimately should, but will do so without having to

[1] Geoff Sayre-McCord spelled this method out to me, after I had explained my (similar) thoughts.

wade through a moral dilemma each time. The exception clause makes it that doctor-assisted suicide is an absolute obligation in certain, but exceptional cases.

A more general form of the Conditional-Substitution method is as follows: for any dilemma which arises from an act-inconsistent moral theory, discern which of the two incompatible actions (in the above example, these actions are: 'harm' (your patient) and 'do not harm' (your patient) presented by the dilemma is the 'proper resolution,' or that action which a moral agent would ultimately perform if given the choice (i.e., 'harm'); then construct a replacement moral theory which is the same as its predecessor, except that it straightforwardly obliges the proper resolution (i.e., 'harm') and straightforwardly prohibits its alternatives (i.e., 'do not harm'); in all other, non-dilemmatic cases, the replacement theory simply concurs with its predecessor (i.e., 'do not harm'). The idea behind utilizing the Conditional-Substitution method is that act-consistent theories are preferable to act-inconsistent theories; therefore, moral dilemmas do not arise when consulting the best moral theories, i.e., the ones that would properly replace our faulty, inconsistent theories. If one still insists that moral dilemmas should persist and, for example, that even morally justified acts of mercy-killing are wrong (*justified* by principles of autonomy but *wrong* according to principles of non-maleficence), then one will be confronted with the arguments of deontic logic against such ideas.

Sayre-McCord (1986) has suggested that all deontic systems have implications for important matters in moral theory. But this is not to say that one must decide moral

matters as deontic logic would have them, for deontic logic is not without its problems. Chisholm (1963), Prior (1954) and Ross (1941) have each suggested that the principles of deontic logic can lead to paradox, so we should not take any deontic principles as uncontroversial. However, controversy or no, the principles of deontic logic gain credibility from the purported analogy between deontic and alethic modalities. Alethic logic is the branch of modal logic that deals with the connections between sentences concerning what is 'necessary,' 'possible,' 'impossible,' etc. Manifestations of this purported analogy include the claim that modalities in alethic logic have exact parallels in deontic logic (McConnell, 1978). So, for example, to deny the agglomeration principle (which, to remind the reader, is encompassed by the notion that: 'ought A & ought B' implies 'ought A&B'), one has to deny an analogue of the agglomeration principle common to all standard forms of modal logic, the principle that 'necessarily P & necessarily Q' implies 'necessarily P & Q.'

The underlying assumption of the view that deontic and alethic logics are analogous[2] is that moral obligation has important similarities to logical necessity. This, surely, is an assumption with which Kant felt comfortable, since he thought that all actions fall into the exclusive and exhaustive categories of 'morally necessary,' 'morally impossible,' or 'morally indifferent.' For Kant, 'a conflict of duties and obligations is

[2] It is important to note that even the apparent analogy between deontic and alethic logic is not perfect. Hughes and Cresswell (1972) claim that although 'it is necessary that *p*' entails '*p*,' and '*p*' entails 'it is possible that p,' these entailments have no obvious counterparts in deontic logic. Further doubts about this analogy have been put forth by von Wright in his later works (1983).

inconceivable'[3]; insofar as grounds for obligation can conflict, at least one of these grounds is insufficient and not grounds for an actual obligation; thus genuine moral dilemmas do not exist.

While cognitivism, deontic logic, and even Kant himself require that our moral codes be free of dilemmas, and while methods such as the Conditional-Substitution argument provide a schematic for constructing dilemma-free moral codes, many moral theorists resist such a program. But even if the controversial nature of deontic logic gives us room to believe that we do not *need* to move away from act-inconsistent moral theories, why is there resistance to using Conditional-Substitution and doing so regardless? The answer lies in examining the driving force behind the Conditional-Substitution argument: the notion that theories that avoid act-inconsistency and do not allow for dilemmas are preferable. This notion is controversial because there are at least three main lines of argument that favor making room for genuine moral dilemmas: the argument from moral sentiment, the argument from a plurality of values, and the argument from single-value conflicts. I will now address each of these arguments in turn.

FOR:

---

[3] See Kant, 'Moral Duties' in Gowan: 1987, p. 39. It must be noted that Kant distinguished between *perfect* and *imperfect* duties. Perfect (or narrow) duties categorically prescribe or prohibit specific kinds of actions. Imperfect (or wide) duties prescribe an unspecified pursuit of ends. These imperfect duties, argues Kant, still must not conflict in the sense that pursuing one precludes the pursuit of the other.

When most people are faced with an apparent dilemma, they find that it cannot be cleanly resolved. This finding is often used as data for a *reductio* against any argument that would have it that genuine moral obligations never conflict. To understand this reasoning, consider the hypothetical code of medical ethics that states 'Never harm your patients, except when respecting their autonomy justifies mercy killing.' A doctor following this code might perform euthanasia on a patient who was as good a candidate for euthanasia as anyone could be, and yet this doctor might still feel a sense of wrongdoing, albeit justified wrongdoing. Insofar as a feeling of wrongdoing indicates an actual wrongdoing, such feelings would cast doubt on the correctness of the code of ethics being followed. Since the code was followed entirely, any wrongdoing would indicate that the rules of the code are not the correct moral rules. Maybe there really is no clean way out of situations where a doctor is asked to kill a patient, even when that doctor has already fully deliberated about what ultimately is to be done in such situations. It has been said that 'ours is a dirty world,' full of tough cases. Maybe sometimes we cannot attain all that we (morally) value; and a code that purports to allow us to do so is unrealistic and inaccurate

The idea that we must account for the feelings of wrongdoing that one has when faced with a dilemma and takes one course out of it is implied by what has become known as the 'argument from moral sentiment.' This argument suggests that we must allow for genuine moral dilemmas so as to account for moral sentiments such as guilt and remorse. Along these lines, Bernard Williams (1973) argues that moral

conflicts are more like conflicts of desires than conflicts of belief. Unlike the way in which a belief can completely give way to an incompatible belief, when a moral conflict is resolved (as when a conflict of desires is resolved), there is a 'remainder,' or a feeling that there lingers a duty not fulfilled. According to Williams, this shows that we do not readily abandon overridden 'oughts.'

To feel the weight of this argument for 'remainders,' reconsider the scenario in which a doctor was asked, even begged to perform euthanasia. If the patient were to miraculously recover and all pleas for death ceased, then the doctor would no longer have a duty to administer a lethal injection. On whatever grounds such a duty was incurred, surely this duty would be canceled upon the patient's full recovery. The doctor might not even need to address the patient on the issue of complying with the original pleas and surely would feel no regret on this matter. Now imagine a case where the patient's condition does not change but the doctor reasons that the requirement to be non-maleficent 'properly' prevails over the requirement to respect a patient's autonomy in such situations. In both cases the doctor decides that 'here, one ought not perform euthanasia,' but surely the two resolutions have a different status. In the latter case, there seemingly remains a latent value in assisting the patient's death, even when the correct resolution of the dilemma has it that one should not do so. When the doctor is deciding what to do in the latter case, autonomy seems to make a genuine call for euthanasia, whereas in the former case, no such call is made. This seems to show that *overridden* duties are somehow different from *canceled* duties. Unlike in cases of

canceled duties where we can wash our hands of the matter, in cases of overridden duties, some explanation, regret, and apology seem to be in order.

As further argument that overridden duties are not canceled, we have Williams' claim that conflicts of moral rules are like conflicts of desires. It certainly seems that the two are analogous, and conflicts of desires certainly seem genuine. For example, consider the desire to eat a second piece of cheesecake and the (let's suppose) conflicting desire to lose weight. When an agent *overrides* a desire to eat cheesecake and refrains from acting on it, this desire seems to linger on as a source of regret and signifies that the agent is not fulfilled. It would seem mistaken to say that, even 'all things considered,' the agent does not have any desire for cheesecake whatsoever. Williams would argue that something would be lost in the analysis if we were to simply say that if one desires to consume large quantities of cheesecake and desires to lose weight then (by agglomeration) one improperly desires to do the impossible: consume large quantities of cheesecake and lose weight.[4] If we then use the Conditional-Substitution method on the conflicting desires, the result would be something like: one only desires to eat cheesecake except in situations like this where one desires weight-loss. But this conditionalized desire is completely fulfilled when one turns down a piece of cheesecake, so why does one keep staring at the desert tray?

---

[4] Notice that while 'ought' implies 'can,' and one is not ever obliged to do the impossible, many might yet think it acceptable to *desire* the impossible, a point to which I shall return much later (see p. 25, below). To this extent, moral conflicts and conflicts of desires might not be as analogous as Williams claims.

The point of the above illustration is to show that, if one is not completely happy, then insofar as this unhappiness is rationally justified, it seems that this conditionalized desire is not the desire that is actually had. Ascribing values (moral or otherwise) to people only if these values disallow conflict appears unrealistic to many philosophers. The agglomeration principle and the Conditional-Substitution method lead us to ascribe sets of desires which fail to account for the sentiment of the agents involved. If such principles and methods are not applicable to conflicts of desires, then maybe they are not to be applied to conflicts of moral values either. If this is the case, moral dilemmas might be genuine.

There are other arguments in favor of genuine moral dilemmas. Resonating with the pluralism of Hegel and British intuitionism, some (Lemmon, Nagel, van Fraassen, and others) argue that there exists a plurality of moral pro-attitudes, and the world is such that these pro-attitudes sometimes oblige contrary actions. Attempts to reduce these values to one supreme value have been denounced as unrealistic, for such a reduction often glosses over essential features of these values. Nagel, van Fraassen and Foot go so far as to suggest that some moral values are incommensurable and so there is no resolution to their conflicts. The notion of incommensurability has received much disdain in the philosophy of science but, for some, remains viable in moral theory. Supposedly, it is viable here because the apparent dilemmas within moral theory seem so perplexing and unresolvable. As an example of such a conflict of values, philosophers often cite the dilemmas described in Shakespeare's *Julius Caesar*. Recall

Brutus' defense of his murdering Caesar: 'it was not that I loved Caesar less, but that I loved Rome more' (act 3, scene 2). Whether the good of friendship should be less valuable than the good of justice is a question that begs for a common ground of comparison. Without knowing of any such grounding, these values appear to be incommensurable.[5]

It has also been argued that whether or not there are a plurality of moral values (commensurable or not), there might exist genuine moral dilemmas. Ruth Barcan Marcus (1980) argues that moral dilemmas can even arise from a single moral value. For example, the requirement to keep all promises allows for situations in which, by no apparent fault of your own, you are in a position that requires your breaking of one promise to keep another. Against this idea, Hare insists that such scenarios demand the consultation of an ultimate moral principle such as the principle of utility, or some other principle that does not allow for moral conflicts. In other words, if a single principle can lead to dilemmas, it is not a morally correct principle to follow. When choosing among incompatible actions, a utilitarian is only obliged to perform the action which maximizes expected utility; if both actions maximize expected utility, the utilitarian is obliged to perform either action, and may 'flip a coin' to decide which action to perform, but is not obliged to perform each. While a utilitarian may be faced with a choice of incompatible actions, it seems that it will never be the case that utilitarian

[5] If it were the case that there are irresolvable dilemmas and thereby no truth-value to some statements about the morally best outcome, moral realism might then be false.

obligations will be attached to both actions and thus no genuine moral dilemmas exist for utilitarians.

Also against the notion of single-value conflicts, Donagan replies that in cases where keeping one promise is incompatible with keeping another, either (1) the agent took unnecessary risks in making both promises and so the conflict is the fault of the agent, not the moral theory, or (2) the promiser is not at fault and is to be somehow released from one of the obligations. The idea behind Donagan's argument is that only an initial wrongdoing can place you in a situation where you are 'damned if you do and damned if you don't.'

There are, however, a few moves to make against the replies of Hare and Donagan, at least as I have briefly presented them. Hare's insistence on an ultimate moral principle might be unrealistic in the same way as attempts to reduce so-called incommensurate values to a single value such as utility: the values that Hare would denounce as mere surrogates for the correct (ultimate) value seem too plausible to discount off-hand. As for my cursory account of Donagan's reply, the notion that one of the obligations must be released does not readily account for the feeling of wrongdoing one has when breaking one of two incompatible promises. Further, against Hare's reply, regardless of whether one can reduce all correct moral values to the value of utility, there exist single-value dilemmas that suggest that even values such as utility-maximizing can yield conflicting duties. While Donagan might be right that one of the obligations must be annulled, when confronted with the dilemmas I am referring to, it

will seem utterly perplexing which obligation that is or how such annulment might go. So, without further ado, I turn to such a dilemma.

Chapter 2: The Deterrence Dilemma ('DD')

Gregory Kavka has suggested (1986: 516-536) that the standards of classical rationality given one's values allow for genuine conflict, even when the desires involved seem rationally permissible to hold. Kavka proposes the infamous Deterrence Dilemma, or 'DD,' as a situation where a single prescription such as 'maximize expected utility' or 'minimize expected harms' can lead to perplexing conflict. To see how the DD might occur, imagine the following situation: A classically rational agent, let us call him 'Max,' leads a nuclear-superpower during a cold-war scenario. Trouble begins when a rival superpower, led by 'Instigators,' threatens Max's country with nuclear attack. Max holds only to one standard, that of 'categorical harm-minimization' and so he always prefers to minimize the harms which are expected to be felt by current inhabitants of the world. To calculate an action's 'expected harms,' Max will determine the number of harms made possible given the performance of the action and multiply this number by the probability that these harms might occur given its performance. Max wants to ensure peace, so he must minimize the likelihood that the Instigators will attack. Max's most effective deterrent against these Instigators is for him to make them believe that he will all-out retaliate against them, should they attack.

The problem is, however, that if Instigators launch their missiles, then Max's country will be doomed; and once Max's country is doomed, Max's retaliation will only kill millions of civilians who might otherwise survive in the attacking country. Categorical harm-minimizers, qua harm-minimizers, cannot rationally prefer to cause

gratuitous harms; therefore, it seems that a harm-minimizer cannot prefer to retaliate in the DD. But if Max believes that retaliation will be an irrational act, then he cannot form classically rational intentions to retaliate and, thus, cannot do what is necessary to minimize harms. This failure is especially problematic since forming such intentions is the only hope of bringing about the highest likelihood of a mutual standoff and thus worldwide peace. Paradoxically, It seems that the DD is a case where the standards of classical rationality get in the way of being able to accomplish what one prefers.

Why cannot Max just bluff, or merely pretend to be a 'would-be retaliator?' The reason is that although bluffing would not commit Max to retaliation, and therefore would not endanger millions of civilians within the Instigators' country, it is stipulated to be unreliable as a strategy of deterrence, and thus does not minimize expected harms. We stipulate that the Instigators have skillful spies, truth-serums, lie-detectors, and keen analytic philosophers, all of which will ensure that only *real* intentions in Max will deter an attack. Max will not only submit to these methods of interrogation, Max might even *suggest* them, since a widespread doubt that he is willing and able to retaliate will only increase the chance that Instigators will attack. In view of this, it seems that a harm-minimizer would not dare to attempt a bluff, and must really intend to retaliate;[6] but as we have seen, this is problematic.

---

[6] Of course, retaliation is conditional upon being attacked, which is a condition Max is trying to avoid.

It would be very troubling to moral theorists if the desire to do what is right could lead to a paradox. Yet this is exactly what the DD might suggest. If minimizing harms is part of the good, then the 'Right/Good Principle' (Kavka 1986) obliges an agent to minimize harms; for this is *'right' and thus is what a good person should do.* But if harm-minimization is good, then the good would demand that Max form the intention to retaliate, which he seems unable to do. Since it seems obvious that 'ought' implies 'can,' something is obviously wrong if Max is unable to do what is good. It is plausible that the standard of harm-minimization is part of the good, yet, as I will further explain, the DD suggests that this standard can create incompatible obligations and thus be act-inconsistent. More problems soon follow. The 'Wrongful Intentions Principle' (Kavka 1986) states that *it is wrong to intend to perform a wrongful act.* Yet the DD suggests that harm-minimizers are obliged to intend to perform what is considered to be an immoral act, namely to cause gratuitous harms. So, if it is wrong to form a retaliatory intention, and if forming such an intention corrupts one's goodness, then one would need to corrupt one's goodness to maintain the peace. But this would be in violation of the 'Virtue Preservation Principle' (Kavka 1986) which states that *one ought never to diminish one's own goodness.* Since it is entirely plausible that minimizing harms plays some part in doing what is right, the above problems illustrate how the DD is of great metaethical import. In fact, since a paradox similar to the DD can be constructed for many types of preferences, e.g., the preference to 'maximize

expected good' ('good' might mean utility, harm-minimization, etc.), the DD may pose problems for many basic principles of morality.

Chapter 3: Previous Solutions to the DD

Philosophers generally agree that there must be a rational way to intend to retaliate, since such an intention would secure the desired deterrent effect. David Gauthier claims (1984: 479-495) that retaliation would be rational if it *expressed a disposition that was rational to adopt.* These 'dispositions', argues Gauthier, put 'constraints' upon what is rational to do. In lay terms, one might say that Gauthier counsels something to the effect of keeping a promise or, in this case, a threat, whether or not you then prefer to do so, ('merely') because you willingly made that promise (threat). I insist that it can never be rational to perform an act that is known to have dispreferred consequences. So if retaliation would be dispreferred by a harm-minimizer, then no dispositions which would lead to retaliation would be rational for her to adopt in the DD. Speaking again in terms of promises, one cannot make real promises while knowing that one will not *keep* them; and no harm-minimizer, *qua* harm-minimizer, would keep a promise or threat to retaliate if doing so would cause gratuitous harms. *If* it is rational to adopt a retaliatory disposition, then it must be that retaliation will *maximize* on the preferences that the intending agent will have when attacked: actions must maximize if they are to count as rational. Since Gauthier's account does not have it that retaliation maximizes on the preferences Max would have when acting, Gauthier fails to show how adopting a retaliatory disposition is rational and thus fails to solve the DD.

Gauthier would reply that it begs the question to demand that actions must maximize if they are to count as rational. After all, Gauthier argues that retaliation is

rational because it follows from a policy that is rational for a harm-minimizer to adopt. Retaliation *does* follow from such a policy, but Kavka, myself and others deny that this is enough to make retaliation rational. Rather than arguing for and against these opposing views of rationality, I will just stipulate that all rational actions maximize and I will later suggest independent grounds for an objection to Gauthier's account.

If Max is unable to intend to retaliate, then maybe Max should step aside for someone or something that can. One might think that Max should build a doomsday machine or relinquish weapons-control to another person, e.g., a trigger-happy Republican, who is disposed to retaliate if attacked. Otherwise, it may remain highly likely that Instigators will initiate an attack. It seems to me, however, that an agent who cared *only* about minimizing harms would not mind killing people who would be killed regardless, because doing so would not increase expected harms. Whether or not such an agent chose to kill the people himself, no fewer people would die, so the choice is moot for a harm-minimizer. Harm-minimizers should be able to do the 'dirty-work' that would otherwise just be done by someone else. In any case, other would-be retaliators might be stipulated as being unavailable in the DD. Maybe Max alone has the ability to launch a retaliatory strike, which consigns the burden of being a would-be retaliator squarely on Max's 'moral' shoulders. If Max is the only possible candidate for being a would-be retaliator, then Max ought to be disposed to retaliate if anyone ought to be so disposed. When Max is or *becomes* a would-be retaliator, Instigators will likely realize this and then it will be likely that no one will get hurt and harms will be minimized.

According to Duncan MacIntosh, Max need not relinquish weapons-control to a would-be retaliator; rather, Max can become one by replacing the preference to always minimize harms with a preference that would permit retaliating when attacked. MacIntosh's theory of 'preference-revision'[7] suggests that in the DD, unconditional harm-minimizers are rationally obliged to revise their preferences so that they come to wholly prefer to retaliate and only *otherwise* prefer to minimize harms, since such a 'revision' of preferences will secure the best chance for a peaceful standoff. Although MacIntosh agrees with Gauthier that retaliation causes gratuitous harms, MacIntosh argues that acting so as to inflict such harms will maximize on the preferences had when acting, since vindictiveness will have become preferable. Since retaliation will maximize on the preferences had when acting, MacIntosh allows his agent to form rational intentions to retaliate and escape the DD.

One might worry that if MacIntosh's agent is attacked, she would re-revise her preferences back to the way they were at the outset of the DD, making it that she would not retaliate. The idea behind this worry is that MacIntosh's agent will realize that the initial preference-revision did not have the desired effect, namely to *prevent* rather than merely 'deter' (minimize the probability of) attack; and so this agent will realize that undoing the previous revision will cause her to refrain from retaliating and will lower expected harms. But MacIntosh can escape this worry, since, even though his agent will realize that a re-revision will lower expected harms, she will no longer

---

[7] MacIntosh (1991), pp. 9-32

desire this result. Once MacIntosh's agent has ceased being a harm minimizer, nothing could convince her to do anything that would lead to non-retaliation; for after revising her preferences, MacIntosh's agent wants nothing more than to retaliate if attacked. Even though a re-revision would lower expected harms, and even though the preference to minimize harms is what initiated the preference revision, fewer harms are not now preferred. MacIntosh's main thrust is that rational agents act according to the preferences they have *when acting*. Thus MacIntosh has suggested how an agent's intention to retaliate might be stable and would thus have the desired deterrent effect. If Instigators understood that the person they were threatening had become and will continue to be a retaliator, they will be deterred.

MacIntosh supposes himself to have shown that we are sometimes obliged to change what we want in order to maximize the likelihood of getting what we first wanted. His proposal implies that unconditional harm-minimizers situated in the DD cannot continue to exist purely as harm-minimizers in what they prefer, because they must either replace their harm-minimizing values, or gratuitously endanger the lives of their compatriots by failing to deter attack. Furthermore, MacIntosh believes that even if harm-minimization is the sole good, ceasing to be an unconditional harm-minimizer will not necessarily diminish one's virtue. For in accordance with his theory entitled 'The Mutability of the Good,' (1995) MacIntosh would argue that Max is morally justified in undergoing self-revision, so long as this 'revision' is kept at the minimum amount needed to allow for intending to retaliate if attacked. According to MacIntosh,

any proper standard of the good will track this change and justify its necessary entailments, e.g., the doing of vindictive actions by a former harm-minimizer. According to MacIntosh, the good 'mutates' during the course of the DD, mutating from something which requires unconditional harm-minimization into something which permits consciously causing gratuitous harms in this one instance.

My Objection to Prior "Solutions" to the DD

I do not believe the proposals discussed thus far are adequate, and analyzing the DD in the context of my previous discussion of the arguments for and against genuine moral dilemmas will help explain why. To present the DD in terms of a standard moral dilemma, I shall explain how the rule 'always minimize harms' could be taken as an act-inconsistent rule. Then I shall reconsider the arguments that there are no *genuine* moral dilemmas, and shall assess the DD's authenticity as a moral dilemma against these arguments.

As it has been characterized, the rule 'always minimize harms' is act-inconsistent. To always minimize harms, one must minimize harms 'pre-attack' (deter) and one must minimize harms 'post-attack' (refrain from retaliating). As I shall illustrate, these obligations are incompatible. First, I must make it clear that one might 'deter,' or *minimize the likelihood of attack*, and still not *prevent* attack. So, just as a failure to *stop* an attack does not necessarily indicate a failure to *hinder* such an act, a failure to prevent attack does not necessarily indicate a failure to 'maximally deter.' But not retaliating if attacked *does* indicate a failure to deter. Given that bluffing does not

maximally deter, if Max does not retaliate when attacked, then Max must not have become 'locked-in' to retaliating; and so he must have failed to deter attack, since anything which was known to be less than a sure commitment in him to retaliation would not have maximally deterred. This shows that if Max is attacked yet does not retaliate, then he failed to minimize harms 'pre-attack' and so failed to do what was then rationally obligatory to do, namely, commit to retaliating in the event of attack, so as to deter attack. To summarize, an agent in the DD has only two metaphysically possible actions: (#1) maximally deter, 'pre-attack', and be disposed to retaliate if attacked, or (#2) fail to maximally deter, 'pre-attack', and thus fail to minimize expected harms. This means that if minimizing harms 'post-attack' entails not retaliating then it also means not having minimized harms 'pre-attack.' So, it seems that one can minimize harms pre-attack or post-attack, but not both. The rule which has it that one should minimize harms unconditionally therefore seems to enjoin mutually incompatible actions.

The DD can now be stated in terms of a standard dilemma: for reasons X (to minimize expected harms, pre-attack), one ought to perform act A (ensure that one will retaliate if attacked); but for reasons Y (to minimize expected harms, post-attack) one ought to perform an act incompatible with A (not retaliate if attacked). We have before us the appearance of a genuine moral dilemma containing a single value, the preference to always minimize harms, which is highly plausible as a moral value and yet leads to

perplexing conflict. Thus, the DD can serve as a test-bed for the previous discussion for and against moral dilemmas.

Since it seems that the rule 'always minimize harms,' at least as Gauthier and MacIntosh construe it, is act-inconsistent, one might use the following argument to cast doubt on the moral correctness of such a rule: to always minimize harms, one must do so both pre-attack (ensure that one will retaliate if attacked) and post-attack (not retaliate if attacked); but applying the agglomeration principle results in the claim that *one ought to ensure that one will retaliate if attacked and not retaliate if attacked*; such an obligation is impossible to fulfill and, since 'ought' implies 'can,' must not be a genuine obligation; therefore, the rule 'always minimize harms' must not be a correct moral rule or must not enjoin such mutually incompatible obligations.

There are several options to take at this point: (1) we can deny that 'ought' implies 'can' and allow that a correct moral rule could oblige impossible actions; (2) we can deny that the agglomeration principle applies to moral obligations and so deny that the rule 'always minimize harms' obliges retaliating and not retaliating; (3) we can concede that the rule 'always minimize harms' is not the correct moral rule and try to argue that this makes the DD metaethically uninteresting; (4) or we can deny that minimizing harms 'pre-attack' is incompatible with minimizing harms 'post-attack' and, illustrating this, solve the DD. I will examine each option in turn, both in light of the arguments for and against genuine moral dilemmas and in light of what Gauthier and MacIntosh have already proposed as solutions to the DD.

First, I will address the denial of 'ought' implies 'can.' This is a denial that neither Gauthier nor MacIntosh would want to embrace. Any morality which required us to do the impossible would seem completely unjust and therefore intuitively incorrect. To see this, one need only imagine a sheriff who stopped unsuspecting pedestrians and demanded that they jump to the moon and back or suffer a huge fine. Such demands seem outrageous and so positing a moral obligation which cannot be fulfilled would seem extremely dubious. While proponents of 'Original Sin' might think that we can be guilty of wrongs that we cannot avoid, many of my readers will allow me to assume that 'ought' implies 'can.'

Secondly, I will address the denial of the agglomeration principle within the DD. One might argue that the agglomeration principle does not apply to moral obligations, claiming that a perfectly rationally permissible preference and/or a perfectly correct moral rule could enjoin incompatible actions. At least, one might demand more evidence for their impropriety than that these rules and/or preferences lead to dilemmas. Before responding to such worries, I want to make it clear that I am now tacitly assuming that 'ought' implies 'can,' and I am addressing my response only to those who accept this assumption. I also want to assume that preferences which are not satisfiable are somehow not strictly rational, although much of what follows can be said without this assumption. This means that those who think that categorical harm-minimization could be a correct moral rule and/or rationally acceptable standard must think that it does not enjoin *impossible* actions, even if they believe that it enjoins

mutually incompatible actions. Obviously, to hold that an act-inconsistent rule (and/or desire) is still 'possible' to follow (and/or satisfy), one must reject the application of the agglomeration principle of deontic logic to moral rules (and/or rational desires). One way of arguing this point would be to say that while it is impossible to minimize harms *throughout* a DD (both pre and post-attack), it is possible to minimize harms 'pre-attack' and 'post-attack,' much as, given a fork in a road, it is possible to travel both routes, albeit only *separately*; therefore, one might insist, it is possible to always minimize harms, even in a DD. But in what follows, I shall explain that this account of how it is 'possible' to always minimize harms is highly problematic, even aside from the fact that the agglomeration principle prohibits such 'separation' of incompatible obligations.

Let us assume that there is only one moral rule: always minimize harms as per the Gauthier-MacIntosh conception of how to do so. For agents who find themselves in a DD where Instigators attack, following the rule is impossible. I have argued that one can minimize harms pre-attack or post-attack but not both. Agents who are attacked in a DD must fail to follow the rule when they retaliate against attack, or failing that, they must have failed to follow it 'pre-attack,' by not ensuring that they would retaliate if attacked. Either way, agents will fail to follow the rule. It is not interesting that an agent 'can,' in some sense, minimize harms at each separate instance in a DD. For the rule demands following the rule *throughout* a DD. An agent *can* minimize harms pre-attack, and that agent *can* minimize harms post-attack, providing we assess this

possibility out of context with what has already transpired; but no one in a DD where instigators attack can 'always minimize harms.' To judge the real possibility of an action, one must do so in context with what has already transpired. And, post-attack, what has already transpired is that Max has ensured that he will retaliate and cause what Gauthier and MacIntosh think are gratuitous harms.

To summarize, there are only two horns of the DD: (1) fail to minimize harms pre-attack; or (2) fail to minimize harms post-attack; and each horn of the dilemma includes an instance of failing to do what Gauthier and MacIntosh think is necessary to minimize harms; to 'always minimize harms' one must never fail to minimize harms, but this is impossible to accomplish in a DD where Instigators attack. So, it seems that I do not need the agglomeration principle, *per se*, to argue that the rule 'always minimize harms' enjoins an obligation which no agent can fulfill. The term 'always' which is embedded in the rule makes the scope of the rule such that pre-attack and post-attack obligations are already conjoined; minimizing harms in either case, but not in both cases, is not sufficient for following the rule. Just as one cannot 'unconditionally' wear a green sweater if one cannot wear green on Thursdays, so one cannot unconditionally minimize harms given the possibility of a DD where Instigators attack. So, if 'ought' implies 'can,' then in a DD where Instigators attack, 'always minimize harms' is not the correct moral rule to follow.

Now I wish to suggest that the rule 'always minimize harms,' as it has thus far been construed, is not a rule that a rational agent would accept. Gauthier and

MacIntosh take themselves to have shown that agents are rationally obliged to retaliate if attacked. Gauthier does this by arguing for 'constrained maximization' and claims that it is rational to express retaliatory dispositions, since such dispositions were rational to adopt. MacIntosh does this by arguing for 'preference-revision' and claims that it is rational to act on revised preferences which target retaliation. So, both philosophers think that Max might be rationally obliged to retaliate, causing what they call 'gratuitous' harms. This means that one cannot minimize harms in all conditions and keep one's rationality intact. If this is not enough to cast doubt on the rationality of an agent who would endorse such a rule, there are further problems. As I have argued, not even irrational agents can comply with the rule 'always minimize harms,' since any source of irrationality would be detected and so would preclude the agent from minimizing harms pre-attack. To always minimize harms in a DD where Instigators attack, one must *ensure* that one will retaliate when attacked and yet not retaliate when attacked, and this is impossible for all agents regardless of rationality or preferences.

The rule 'always minimize harms,' whether rational to embrace or not, is impossible to comply with in a DD where Instigators attack and so is not there a correct moral rule. It is plausible that the correct moral rule would be 'always minimize harms except when in a DD where Instigators attack, in which case one should retaliate.' As far as I can see, such a rule can be followed perfectly, even given the possibility of the DD. If following this rule is 'right' then a good person could do so without corrupting her virtue or forming a wrongful intention. The DD does not now

seem to pose a problem for the basic principles of morality, since agents who target the correct moral rules cannot fall prey to the DD.

At this point the reader might not see the difference between my analysis and that of MacIntosh. But reconsider MacIntosh's argument: he attempts to solve the DD by arguing that the 'newly conditionalized' preference, *minimize harms at all times except when in a DD where Instigators attack*, must be adopted; but a preference to do the impossible, which is what minimizing harms unconditionally amounts to, would need to be 'replaced,' not 'revised.' Recall that MacIntosh's 'preference-revision' is supposed to be thought of *as changing what you want in order to maximize the likelihood of getting it*, not as *changing what you want because you could never get it*. For any agent who wants to minimize harms unconditionally, preference-revision will not increase the likelihood of satisfying this desire, for the probability of doing so remains at zero, regardless of what one does or prefers.

Along the same lines, MacIntosh's theory of the mutability of the good is problematic since it was never the case that one was obliged to minimize harms unconditionally, for there always lurked the condition of a DD where Instigators attack as a counterexample to the correctness of the rule 'always minimize harms.' The suggestion that the good has 'mutated' so as to make acceptable a failure to minimize harms (at some point during a DD where Instigators attack) is to suggest that prior to this mutation, it was morally unacceptable to do so; yet it was never morally unacceptable to do the unavoidable and it is unavoidable to fail to minimize harms in a

DD where Instigators attack. Even agents who are never confronted with a DD are not obliged to be disposed to 'always minimize harms' when thrown into a DD, and so the good does not have to mutate to allow for such failures to minimize harms; the good must have allowed such failures all along!

Let me summarize my primary innovations thus far. Given that 'ought' implies 'can,' I have suggested how act-inconsistent rules such as 'always minimize harms' are impossible to follow and therefore not the correct moral rules. I have done this without appealing to the agglomeration principle, *per se*. This said, I have what I believe to be a new proposal for dissolving the DD and other dilemmas which utilize rules that lead to the sort of act-inconsistency which I have attacked. My proposal has implicit counter-arguments against MacIntosh's theory of Preference-Revision and of the Mutability of the Good. Ultimately, however, I contend that a superior proposal is in the making. This proposal involves taking the above-mentioned option (4): we can deny that minimizing harms 'pre-attack' is incompatible with minimizing harms 'post-attack' and, illustrating this, solve the DD.

Chapter 4: My Proposed Solution to the DD

Suppose that an agent is properly called a 'would-be retaliator' if and only if that agent would retaliate if attacked in the DD. Only real 'would-be retaliators' will pass as such, since any who merely pose as such can be stipulated out, so only genuine would-be retaliators minimize expected harms 'pre-attack.' Take any agent who is attacked in the DD: if this agent has minimized expected harms thus far, then this agent is a would-be retaliator, and as such, will retaliate. Post-attack, the only agents who could possibly succeed in minimizing harms unconditionally are ones who have minimized harms thus far; but the only ones who have minimized harms thus far are would-be retaliators. Therefore, if the DD in which Instigators attack is a possible condition, and if unconditional harm-minimization is possible, then 'post-attack' retaliation must here count as minimizing expected harms, since the only agents who have minimized harms thus far will retaliate if attacked. My analysis of the DD has it that either (1) the DD is dissolved as a dilemma of morally correct rules, since it is impossible to always minimize harms; or (2) it is possible to always minimize harms, thus the DD is not dissolved, but it must somehow be that retaliators count as minimizing harms post-attack.

I doubt that the DD entails that the preference for unconditional harm-minimization, the very thing the DD assigns as a preference of rational agents, is an irrational value. I maintain this, even though I am highly sympathetic with arguments that no standard of the good can enjoin mutually incompatible actions. So I contend

that unconditional harm-minimization is *not* act-inconsistent and that one does not have to use arguments against genuine moral dilemmas to *dissolve* the DD. The rule that one ought always to minimize harms is not act-inconsistent in the way that makes abiding by it impossible, because whenever one can choose so as to expect various amounts of harms, one can choose so as to expect the minimal amount; and, if the amount of harms to be expected will remain the same, regardless of what one does, then this fixed amount of harms *is* the minimum amount that one should expect; so, whether the amount of expected harms is variable or not, one can *always* (even in the DD) choose so as to minimize them. All of this suggests that agents can remain unconditional harm-minimizers. As outlined above, this entails that retaliators must count as minimizing harms post-attack when they retaliate.

Possible Objections and Counterexamples to My 'Solution'

What repercussions for action-theory does my proposal have? And what action-theoretic assumptions are required for my 'solution' to be viable? Am I not allowing determinism to creep in, or even worse, assuming it from the outset, when I argue that someone who intended to retaliate *must* retaliate if attacked? It may well be that Max only 'acts' in adopting the intention to retaliate; after that, Max might be 'locked-in' and does not perform an act, *per se*, when retaliating. Maybe Instigators make 'pre-attack' demands that Max make a *final* decision to be a would-be retaliator or not to be one. Once a *final* decision has been made, there are, by the definition of 'final,' no more deliberations or 'acts' to perform. There are no unforeseen options that would

require that further decision-making occur in the DD. One could say that it is the 'pre-attack' choice to be a harm-minimizer that restricts one's 'post-attack' options. Max may be rationally obliged to become so much like a doomsday machine that his freedom of choice must be left behind.

My argument is not committed to the view that retaliation is not an act. Max's decision to retaliate must be final, but it is a decision on how he will act. Max must become somewhat like a doomsday machine, but only in the sense that he must *act* like one; when Max intends to retaliate, Max becomes disposed to *choose* to retaliate. Even if retaliation is an act, there is a significant sense in which Max *had* to act as he did. In a DD, for Max to refrain from retaliating when attacked, it would have to be that he had refrained from forming retaliatory intentions. But how could this have transpired? Consider that Max was capable of forming a retaliatory intention, he believed the intention's presence would minimize harms, he was rational, and he preferred to minimize harms; one of these four things would need to be different if Max were to refrain from initially forming a retaliatory intention. I grant that there are things that might allow Max to avoid forming a retaliatory intention, e.g., a change in the belief that the intention would deter, a revision of the preference to minimize harms, or a bout of irrationality which would allow Max to do other than what was rationally obligatory. But, I insist that none of these things are generally thought to be under an agent's power to bring forth. Since no ingredient required for the absence of a retaliatory intention is had in the DD, and since Max could not bring forth such ingredients, the

absence of a retaliatory intention remains impossible. In a normal DD, wherein Instigators do attack, just as the intention to retaliate is necessary, so is retaliation.

I am not now denying MacIntosh's claim that preference-revision is *sometimes* rationally obligatory. But while a revision in the direction of Max becoming vindictive might be possible, a revision which would remove this vindictiveness is not. As earlier suggested, it is for the very reason that 'post-attack' revisions are rationally impossible that MacIntosh first suggests 'preference-revision' as a way for Max to deter attack: 're-revisions' are known to be rationally impermissible and Instigators fully believe that attacking Max will be suicidal. Therefore, since Max is (and always was) a would-be retaliator, the minimum harms actually possible for him will depend solely upon whether or not Instigators decide to attack. In fact, since Max always was a would-be retaliator, then, strictly speaking, he does not become disposed to retaliate upon forming retaliatory intentions, he rather has been so disposed upon initial acquisition of the desire to minimize harms.

Regardless of the correct theory of action, my argument still stands: denying that a retaliator minimizes harms implies (via my earlier argument) that the DD is dissolved as a problem for ethics. But if it must be possible to minimize harms unconditionally, one cannot deny that straightforward retaliators minimize harms whether they do so by choice or not. One might say that conditional retaliation is a logical consequence of being a harm-minimizer, so if one ought to be a conditional harm-minimizer, one ought to conditionally retaliate.

How would one characterize the *choice* to retaliate? If immediately after a retaliatory strike has been launched, someone asks Max why he did not refrain from retaliating, then Max can simply reply, 'As we all knew that it would be, it was rational for me to retaliate, and so I did. Sadly, the Instigators have rejected the chance for peace that I have created and they have doomed us all. For it was the Instigators who launched their missiles towards us and they who have made it so that I would choose to return the favor. The Instigators will go down in history as having caused their own deaths, for they knew how I would react when they attacked me. I cannot now prefer to harm fewer people, for how could I now choose differently? If I had been the type of person who would not retaliate, then we would not even be alive today for you to support or second-guess this decision of mine, and we would have missed the opportunity for a mutual peace. And furthermore, if we had died without returning the favor, or if we had been unexpectedly spared despite our cowardice, it would have been after doing something that I could never do; for I would never put our own lives in gratuitous jeopardy by not trying to deter those who threaten to attack.'

There are two possible responses to my characterization of the choice to retaliate: (1) it seems that my above description is much alike Gauthier's constrained maximization in that retaliation does not seem to express the preference to minimize harms even if it expresses a disposition it was rational to adopt; (2) to the extent that I can make retaliation maximize on the preferences had when acting, something I think MacIntosh rightly demands, I may have to concede that Max must have changed his

preferences if he chooses to retaliate. In other words, one might see that it is necessary that Max has the preference to retaliate, might not blame him for acting on this preference, but deny that this preference is that of a harm-minimizer; if Max has a choice in the matter about how many harms he inflicts post-attack, and if Max prefers to inflict the higher of the two amounts of harms, it surely seems that he is not a harm-minimizer. Even though Max's post-attack preferences did not come by way of 'Preference-Revision,' but by some sort of 'replacement,' many will insist that Max's are not the same preferences held at the outset of the DD. But given that it is possible to 'minimize harms' post-attack, I now argue that Max might succeed in following the rule 'minimize harms' even when another agent would have chosen to harm fewer people.

Consider that if some agent had the ability to disarm everyone's missiles in mid-air, such an agent might choose so as to ensure that none are harmed post-attack. Such an agent harms fewer people than Max does. But since this 'disarming' ability was absent in Max, we must set the amount of harms that are properly called 'the minimum (possible in a DD)' higher than we might otherwise have to. Harm-minimization, like other candidates for the good, might be 'context-sensitive,' in the sense that the 'minimum' amount of harms possible in one situation (the situation composed of the dilemma, and of the particular agent facing it), might be different in another. We might say of an agent who disarmed all of the Instigators' missiles that both she and Max minimized harms given their different situations. In fact, some agents might harm fewer

people than Max and yet fail to minimize harms, if it is the case that their situation would have allowed them to harm fewer still.

The objection can be made that while different agents with different abilities make for different situations, agents who merely have different preferences do not make for different situations. For example, I would not want to call a war monger a 'harm-minimizer' simply because, *given his preferences*, he is in a situation where he cannot harm fewer people than he does. This is problematic for me, since I have argued something quite similar when claiming that Max, an agent whom I claim prefers only to minimize harms, has the choice to harm fewer people than he does, but, given *his* preferences, rationally *cannot*. But what may differentiate the cases between Max and a war-monger is that Max's preferences are those which, at the outset of any situation, are most expected to minimize harms. Max is an agent who clearly is aiming at what we think to be the right standard, and he is doing everything that we could expect him to do to ensure his own compliance with this standard. Anyone who harms fewer people than Max harms is in less difficult situations or merely beats the odds. Although I have left room to maneuver against me here, I suggest that the correct ethical theory must grant Max the status of 'morally-good agent,' since he alone does what a rational person would do given a desire to target what is right.

If we grant that the correct moral theory has it that moral agents are fully rational and target the correct moral rules, at least at the outset of moral dilemmas, then I can make one last attempt at persuading the reader that retaliators minimize

harms. Given Max is a rational agent who prefers to minimize harms pre-attack, it is also a given that he will want to retaliate should the time come when he is attacked. Keeping this in mind, consider the following brief thought-experiment. Imagine that I told you that at the end of a hallway there were two lollipops, one large and one small. Imagine also that I convinced you that *if and when* you got to the end of the hallway, you would prefer large lollipops over small ones. To convince you of this, perhaps a lie-detector, a psychologist, and a large guillotine are made visible about half-way down the corridor. Would it not be odd for you to boast that you could pick the small lollipop? Likewise, would it not be odd if we boasted that we could minimize harms pre-attack and then choose to refrain from retaliating if attacked? So it seems that, until we can fully explain how it is to be done, we cannot make post-attack insistences that Max be a non-retaliator.

Let me now sum up my position on the DD and further illustrate a crucial distinction between Gauthier's position and mine. Take any rational agent who values nothing but harm-minimization. Such an agent will prefer to minimize harms wherever possible, given what has already transpired. Minimizing harms 'wherever possible, given what has already transpired' will be equivalent to either (1) doing so unconditionally, since it is possible to always minimize harms, even in the context of what has already transpired in a DD where Instigators attack; or (2) doing so in all conditions except in a DD where Instigators attack, since harm-minimization is not there possible for someone who minimized expected harms pre-attack. The point of all

this is that whether Max holds an unconditional or a conditional preference to minimize harms, contrary to Gauthier's account, retaliation will maximize on this preference. Given that he is maximizing on his only preference, Max will feel no regret in killing millions of people.

In response to my proposal, Williams might worry that something has gone awry when a harm-minimizer kills millions of people without regret. It is to this response that I will now turn.

Chapter 5: Implications for Dilemmas and Regret

Sentiments of regret seem to signify one's lack of fulfillment. Sometimes, as in the case of so-called moral dilemmas, it seems that even agents who do their best to follow what are the correct moral rules come out of the dilemma feeling regretful. In some situations, there seems to be no 'clean' way out. If agents target only the correct moral values and can still feel regret, and if regret signifies a lack of fulfillment, it seems that doing our best might not be enough to fulfill the correct moral obligations.

If I am to argue against the possibility of genuine moral dilemmas, I must explain why the abundance of regret in the world is not sufficient evidence for such dilemmas. To do this, I will suggest that regret is had more often than it needs to be and then I will suggest that the regret that lingers in the mind of a rational agent might be explained without positing unfulfilled moral duties. To do this, I begin with the following thought experiment.

Imagine that one day you are walking along and you see a five- and a ten-dollar bill lying on the curb. You are not the type of person to pass up 'easy money,' so you reach for both bills. Suddenly a swirl of wind lifts the money away from you and blows it into the busy street. You only have the opportunity to grab one of the bills, so you grab the 'ten' and surrender the 'five' to the traffic. You are late for work and cannot afford to search for a five-dollar bill that is probably in someone else's pocket by now. So you continue on your way, ten dollars richer. In this scenario, I have presented two incompatible actions (grabbing the five dollars vs. grabbing the ten dollars) which stem

from the desire to never pass up easy money, and, although it is clear which bill you should grab, it is not clear how you should feel about the bill that you have left behind.

It has been said that psychology requires us to have certain sentiments when taking on one horn of a particular dilemma.[8] For one thing, such sentiments might serve to ward off future conflicts. I grant that the taste of painful sentiment that arises from our mistakes trains us to avoid making similar mistakes in the future, but taking the least painful horn of a dilemma is no mistake at all. Being 'no mistake at all,' taking on the lesser of two so-called evils is not something to be regretful about nor is it an indicator of future problems. Take, for example, a person who feels absolutely no regret over the loss of the five-dollar bill when grabbing it would have entailed losing a ten-dollar bill. There is no reason to think that such a person would be careless about money in the future, or choose to be in situations where one could get either a five- or a ten-dollar bill rather than situations where one could get both. My point is that since this person has maximized the easy money obtained thus far, there is nothing to say that this person will fail to do so in the future. If the person needlessly passed over easy money and felt no regret, then *this* might indicate an indifference towards easy money; maybe such a person would needlessly allow conflicts of 'easy money' to arise, not trying to avoid situations in which grabbing one bill conflicted with grabbing another when both were otherwise readily available. But none of this has thus far occurred.

---

[8] Susan Sherwin made this point to me. Likewise, Ruth Barcan Marcus, in 'Moral Dilemmas and Consistency,' argues that the guilt and feelings of wrong doing help people avoid future dilemmas.

One might say about the 'easy money dilemma' that no regrets are to be had; for you did not have a rational desire to grab the five-dollar bill since you knew that you could only grab a single bill and wanted that bill to be worth as much as possible. What you wanted to do, one might insist, was to grab as much easy money as you could; and you have done exactly that. Any regret that arises over the loss of the five-dollar bill is highly suspect and not necessarily the sort of thing that one should account for in a theory about what standards one ought (morally and/or rationally) to strive for.

How do these considerations pan out in the case of conflicts of moral obligations? Even though the principle of respecting autonomy might justify a doctor in 'harming' a patient by assisting in that patient's suicide, many will claim that a doctor who could do such harm without regret is likely to be careless when such matters arise in the future. But as in the case of the five- and ten-dollar bills, I argue that the doctor in question has maximized the good thus far, and so might reliably do so in the future. But even if this doctor is efficient and can be relied upon to maximize the good, is the fact that he or she does not feel regret a sign of some other moral defect? Even if it were the case that the correct rule is 'Be non-maleficent, except when respecting autonomy requires hurting someone,' then would the doctor necessarily care whether autonomy and non-maleficence ever conflicted? We care about such things greatly. There certainly seems to be something *prima facie* desirable about a world in which no one is ever harmed. Does this not suggest that non-maleficence is a good thing, even

beyond the extent to which it is compatible with achieving other goods? And does this not also suggest that our doctor is somehow lacking in moral qualities?

To answer these questions, I will consider the utilitarian's single moral obligation: always maximize expected utility. A utilitarian can admit that it would be even better to produce more utility than what the given situation dictates is 'the maximum,' but would deny that there is now a duty to do so. Likewise our doctor can want that non-malificence never needs sacrificing to promote patient autonomy; for this would be a natural result of wanting minimal harms. No amount of utility beyond that which *can* be brought about, nor any smaller amount of utility which must be sacrificed in order to bring about the maximum amount is a 'good' in the sense of being something which we are obliged to actualize. The utilitarian feels no regrets about sacrificing lower utility for higher utility. Likewise, our doctor need not feel the kind of regret which signifies an unfulfilled moral duty in order to avoid careless mistakes or to wish that the world were a better place.

Returning to the 'easy-money' scenario, imagine that you have grabbed the ten-dollar bill in lieu of the 'five' and, only a few minutes later, the wind deposits that same five-dollar bill on a nearby curb. You reach for the bill, but before you can grab it, a teenager on Rollerblades snatches it up and bolts in the opposite direction. You'll never catch the teenager, and even if you could, you do not have a legitimate claim to the money; so you hurry on towards your prime source of easy money, namely your work. The question now becomes: do you feel the same sentiments over the 'loss' of the five-

dollar bill each time it has presented itself? In the first instance, all that kept you from grabbing the five was your desire to grab the ten, in the second, contingencies about the world made it impossible to grab the five, regardless of your desires. So, how do you feel about each instance of failing to grab the five dollars?

What I want to suggest is that the regret over a so-called 'desire' that was overridden may be just as specious as the regret felt when 'desires' simply cannot be fulfilled. So, maybe one should feel the same way when five dollars is taken away from you as when you choose to forego it. There is, however, an argument to make against the analogy between *overridden* 'desires' (and/or 'duties') and those which simply are impossible to fulfill. Unlike in cases in which a desire simply cannot be satisfied, it seems that in the case of overridden 'desires' and or 'duties,' *it was possible to fulfill them*. It was possible, for example to grab the five dollar bill on the first occasion that we saw it, but not the second. Likewise, it is possible to avoid harming others when all that stops you is your desire to respect autonomy, but not when you are on an unavoidable collision-course with a small family. So, it may be the sense in which overridden 'desires' are possible to fulfill that makes it seem regrettable when they are left unsatisfied; for it is the agent's overriding desire that keeps the 'desire' unfulfilled when *overridden*, rather than the world. To this extent, the agent may be seen as playing a larger role in the dissatisfaction of the overridden desire and thus may be seen as more responsible and so more troubled by what has transpired.

Many would also argue that we would not want to be able to go through tough situations without feeling some regret; for only a moral monster would feel no regret when taking one horn of a particularly hard dilemma. Reconsider our old friend, Max, who I contend can kill millions of people without a single unsatisfied preference. In fact, I argue that Max would be happy to do so. Does this not signify that there is something very wrong with the state of mind that I am counseling? Not necessarily. Even those who disagree with my analysis of the DD would likely agree that, if presented with the option of killing either one million or two million people, Max would kill one million people. Since killing as few as possible completely maximizes on Max's desires, he can do so without regret. If presented the same options, I doubt that *I* could remain so cool-headed, but my preferences are not those of Max. Notice that Max will react differently than you or I might, not because I have misinterpreted how Max's preferences are to be satisfied, but because Max has a very different set of preferences than you or I. Max is a highly idealized agent who *only* wants to minimize harms, and I have argued that in a DD where Instigators attack, killing millions of people accomplishes that. The question I posed about the DD was not whether Max could kill a million people without regret, rather it was whether the one million deaths in a DD can be properly called 'the minimum.' The answer to the latter question ('yes') will lead us to the answer to the former ('yes'). Any thought that Max 'ought' (morally and/or rationally) have regret over the deaths he has caused implies that either Max

ought to have different preferences or Max has failed to maximize on those he has; but neither condition has been shown to be the case.

Moving away from Max, let us now consider an agent with a more 'human' set of preferences: the doctor who is asked to assist in a patient's suicide. Suppose that this doctor treats overridden duties as if they were completely canceled, and believes that wherever respecting autonomy precludes being non-maleficent, one has absolutely no duty to be non-maleficent. If regret stems from unfulfilled obligations, such a doctor could sometimes harm patients without regret. But before judging the doctor too quickly, consider that he or she could be heartbroken about the situation, even while maintaining that there exists no duty to be non-maleficent in such cases. This is apparent when we consider the feelings that one would have if one accidentally drove one's car over a young family and killed them all. There is no duty to avoid unavoidable harms, yet strong sentiments seem appropriate when they occur. My point is that we might require agents to feel strong sentiment without requiring that the sentiment entails that some duty has been left undone. Whether or not the doctor feels upset about the situation does not entail that he or she is following the wrong set of moral principles. Granted, it might be premature to say that all regret is merely a pining over the impossible and a denial of the unavoidable, but this is probably the explanation more often than we suppose.

While many might insist that it is, in some isolated and non-agglomerated sense, 'always possible' to minimize harms, or keep promises, or respect the principles of

medical ethics, etc., this is not the possibility that is required to keep these rules viable as part of the good. We need more than that it is *always possible to minimize harms*, it must be possible to *always minimize harms*; in the first case, we might argue about whether dilemmas preserve this possibility, but in the second case, dilemmas preclude it. If I am correct, we can never fulfill a rule that is act-inconsistent, since the rule is not complied with unless both (incompatible) actions are performed; this is impossible, so we are not morally obliged to do it.

Chapter 6: Concluding Remarks:

So it seems that a large source of regret is our mistakenly measuring ourselves against what we think are unactualized 'possibilities' but which turn out to be impossibilities (i.e., not retaliating if attacked); we think that we can have our cake and eat it too and we feel guilty when we fail to do so.

When we hear that Max has retaliated, our first reaction is to insist that he could have done better, for 'all he had to do was refrain from retaliating.' But once we put ourselves in Max's shoes, we have no idea how what we have insisted upon is to be accomplished. We mistakenly hold people responsible for failing to meet certain standards even though we are unable to say exactly how meeting that standard might go. To say that Max can just 'not retaliate' is too quick; for how is he to do this given what has already transpired by the time we demand this of him?

It may be that the mere *appearance* of moral conflicts is all that we need to fuel regret. Maybe the standards to which we aspire are, unbeknownst to us, unreasonable. We have been taught various principles at various times from various people who have various moral theories. So it might be no surprise that the system that we have had cut-and-pasted into our minds is incomplete, contradictory and, as a whole, false. But we have not sorted that out yet, so we feel regret when we do what our incorrect theories deem is 'wrong.' For example, we might feel horrible feelings of wrongdoing after losing our virginity, but we might later find that our sentiments of guilt (sexual or otherwise) stem from something other than a recognition of true moral principles. If

regret is ever justified, I would venture to guess that there are also many cases where we feel much less regret than we ought to. An example of something that we might not regret enough is our callously letting distant others die of starvation while we commonly enjoy an abundance of food. Our present moral intuitions might serve as a rough guide to right and wrong, but we probably have a long way to go before our sentiments accurately reflect an enlightened and true morality. The fact that we sometimes have done the best that we could and still feel unfullfilled should tell us that we are asking too much of ourselves.

Bibliography:

Chisholm, R. M. "Contrary-to-Duty Imperatives and Deontic Logic." *Analysis* 24 (1963): 33-36.

Copp, D. and Zimmerman, M., eds. Morality, Reason, and Truth (Tottowa, NJ: Rowman and Allenheld, 1984).

Donagan, A. Choice: The Essential Element in Human Action (New York: Routledge & Kegan Paul, 1987): 94-112.

Gauthier, D. "Deterrence, Maximization, and Rationality." *Ethics* 94 (1984): 479-495.

_____. Morals By Agreement (Oxford: Clarendon, 1986), section VI. "Compliance: Maximization Constrained": 157-189.

Gowans, C. W., Moral Dilemmas. (New York: Oxford University Press, 1987).

Harding, C. G. "Intention, Contradiction, and the Recognition of Dilemmas" in Harding, C. G. Moral Dilemmas (Chicago: Precedent, 1985): 43-56.

Hare, R. M., Moral Thinking. (Oxford: Clarendon Press, 1981): 25-43.

Hughes, G. E. and Cresswell, M. J. An Introduction to Modal Logic. (London: Methuen, 1972).

Jackson, F. "Davidson on Moral Conflict" in LePore and McLaughlin. Ed. Actions and Events: Perspectives on the Philosophy of Donald Davidson (New York: Blackwell, 1985): 104-115.

Kavka, G. "Some Paradoxes of Deterrence." In John Perry and Michael Brattman, Eds. Introduction to Philosophy: Classical and Contemporary Readings (New York: Oxford University Press, 1986): 516-536. Originally published in *The Journal of Philosophy* 75 (1971): 285-302.

_____. "The Reconciliation Project" in Copp and Zimmerman, eds. (1984): 297-319.

Lemmon, E. J. "Moral Dilemmas." *Philosophical Review* 70 (1962): 139-58.

MacIntosh, D. "The Mutability of the Good." Read by him at Dalhousie University, 1995.

_____. "Preference-Revision and the Paradoxes of Instrumental Rationality." *The Canadian Journal of Philosophy* 22 #4 (December 1992): 503-530.

_____. "Retaliation Rationalized: Gauthier's Solution to the Deterrence Dilemma." *Pacific Philosophical Quarterly* 72 (1991): 9-32.

Mally, E. "*Grundgesetze des Sollens, Elemente der Logik des Willens.*" (Graz: Leuschner, 1926). Reprinted in Logische Schriften: Grosses Logikfragment-Grundgesetze des Sollens. Eds. Karl Wolf and Paul Weingartner (Dordrecht: Reidel, 1971): 227-324.

Marcus, R. B. "Moral Dilemmas and Consistency." *Journal of Philosophy* 77 (1980): 121-36.

McConnell, T. C. "Moral Dilemmas and Consistency in Ethics." *Canadian Journal of Philosophy* 8 (1975): 269-87.

Narveson, J. "Reason in Ethics—or Reason versus Ethics?" in Copp and Zimmerman, eds. (1984): 228-250.

Neilson, K. "Must the Immoralist Act Contrary to Reason?" in Copp and Zimmerman, eds. (1984): 212-227.

Oates, W. J. The Stoic and Epicurean Philosophers (New York: Random House, 1940): 223-468.

Peacocke, C. "Intention and Akrasia" in Bruce Vermazen and Jaakko Hintikka, eds. Essays on Davidson: Actions and Events (Oxford: Clarendon, 1985): 51-74.

Penner, T. "Plato and Davidson: Parts of the Soul and Weakness of the Will," in David Copp, ed. Canadian Philosophers (Calgary: CJP, 1990): 35-74.

Prior, A. N. 1954. "The Paradoxes of Derived Obligation." *Mind* 63 (1954): 64-65.

Ross, A. "Imperatives and Logic." *Theoria* 7 (1941):53-71.

Sayre-McCord, G. "Deontic Logic and the Priority of Moral Theory." *Nous* 20 (1986): 179-97.

Shick, F. Understanding Action (Cambridge, MA: Cambridge University Press, 1991): 42-45, 110-120.

Sinnott-Armostrong, W., Moral Dilemmas. (Oxford: Basil Blackwell, 1988): 72-109; 169-188.

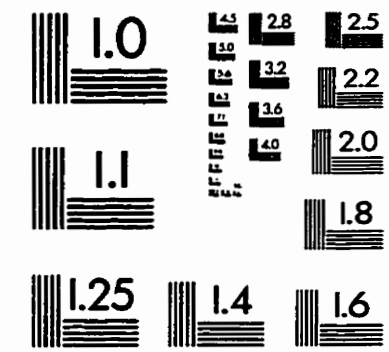
Von Fraassen, B. C. "Values and the Heart's Command." *Journal of Philosophy* 70 (1973): 5-19.

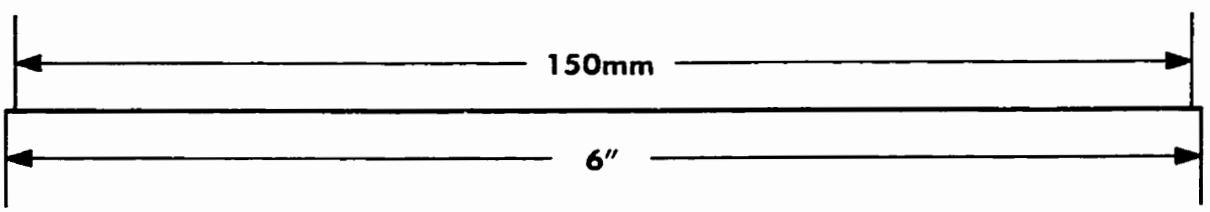
Von Wright, G. H. "Deontic Logic." *Mind* 60 (1951): 1-15.

_____. "On the Logic of Norms and Actions." Practical Reason. Vol. 1 of Philosophical Papers. (Ithaca, NY: Cornell University Press, 1983): 100-29.

Williams, B. "Ethical Consistency." Proceedings of the Aristotelian Society, Supplementary Vol. 39 (1965): 103-24.

_____. Problems of the Self: Philosophical Papers 1956-1972. (Cambridge: Cambridge University Press, 1973).





APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989